ABSTRACT

Representative embodiments are directed to systems and methods for processing training data for a statistical classification application. In one embodiment, confidence values are calculated for training data elements to identify the probabilities of the training data elements belonging to identified classes. An interactive scatter plot is generated using the calculated confidence values. The scatter plot visually indicates the confidence values of points in the scatter plot. Accordingly, the user is able to identify potentially misclassified training data elements. The user may select training data elements from the scatter plot and reclassify training elements as appropriate. Upon reclassification, the confidence values may be recalculated and the scatter plot revised.

25377692.1